Article

SOCIETY FOR THE TEACHING OF PSYCHOLOGY

Teaching of Psychology
1-6

DOI: 10.1177/0098628316677492
top.sagepub.com
SAGE

Ecological Validity of the Testing Effect: The Use of Daily Quizzes in Introductory Psychology

W. Robert Batsell Jr.[1], Jennifer L. Perry[1], Elizabeth Hanley[1], and Autumn B. Hostetter[1]

Abstract

The testing effect is the enhanced retention of learned information by individuals who have studied and completed a test over the material relative to individuals who have only studied the material. Although numerous laboratory studies and simulated classroom studies have provided evidence of the testing effect, data from a natural class setting with motivated students are scant. The present two-class quasi-experiment explored the external validity of the testing effect in the Introductory Psychology classroom. The control class studied assigned chapters from the textbook whereas the quiz class studied chapters and completed daily quizzes on those readings. Subsequently, both classes completed exams over this textbook information. The quiz class scored significantly higher than the control class on these test questions about the textbook information; these differences were significant both when the test questions were the same as the quiz questions and when they were new, related questions from the textbook. These data suggest the use of daily quizzes to embed the testing effect into the Introductory Psychology classroom can improve student learning.

Keywords

Testing Effect, Quizzes, Spacing Effect, Effortful Recall, Introductory Psychology

Psychology instructors should attend to the latest findings from basic research regarding learning and memory and then determine whether these findings can be incorporated within efficacious teaching techniques. One recent research finding that appears to have direct application to the classroom is the testing effect (e.g., Karpicke & Roediger, 2008; Roediger & Karpicke, 2006), in which participants show superior retention for information after being tested on that information. For example, Roediger and Karpicke reported that participants given the opportunity to study a passage and be tested on that passage (the testing group) demonstrated better retention of the passage after a delay (2 or 7 days) compared to participants who studied the passage twice. Significant testing effects are not limited to expository-based or text-based materials, as they have been reported with other materials, such as visuospatial tests (e.g., Carpenter & Kelly, 2012), map learning (e.g., Carpenter & Pashler, 2007), videos of lectures about art history (Butler & Roediger, 2007), narrated animation (Johnson & Mayer, 2009), symbol–word pairs (e.g., Coppens, Verkoeijen, & Rikers, 2011), and function learning (e.g., Kang, McDaniel, & Pashler, 2011). Not only has the testing effect been shown with a wide range of test materials, it has also been shown with a wide range of test types, ranging from fill-in-the-blank tests (Hinze & Wiley, 2011) to short answer quizzes (Kang, McDermott, & Roediger, 2007) to open-book tests (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008). Importantly, the benefits of the testing effect are not due simply to differential exposure to the test questions. For example, Chan (2010) has shown the presence of transfer or *retrieval-induced facilitation*, which is the phenomenon that the testing effect enhances retention not only of the material on the initial test, but it also increases retention of nontested material from the original passage (see also Rohrer, Taylor, & Sholar, 2010). In sum, a number of laboratory studies suggest that testing produces enduring memory with a range of stimulus materials.

Many investigators have attempted to determine whether the robust testing effects seen in the lab can also be achieved in the classroom (e.g., Butler & Roediger, 2007; Cranney, Ahn, McKinnon, Morris, & Watts, 2009; Hattikudur & Postle, 2011; Khanna, 2015; McDaniel, Agarwal, Huesler, McDermott, & Roediger, 2011; McDaniel, Anderson, Derbish, & Morrisette, 2007; Roediger, Agarwal, McDaniel, & McDermott, 2011), but to date, evidence demonstrating significant effects under

[1] Kalamazoo College, Kalamazoo, MI, USA

Corresponding Author:
W. Robert Batsell Jr., Kalamazoo College, 1200 Academy St., Kalamazoo, MI 49006, USA.
Email: robert.batsell@kzoo.edu

real-life conditions is sparse. The primary concerns with extending these studies' findings to the college classroom are related to population validity and ecological validity. Pertaining to population validity, researchers have shown a significant testing effect in an eighth-grade science class (McDaniel et al., 2011) or a sixth-grade social studies class (Roediger et al., 2011), yet these results may not be replicable with college students and their advanced approaches to studying test materials. Regarding ecological validity, studies have been conducted predominantly in simulated classroom settings (e.g., Butler & Roediger, 2007). When they have been conducted in actual classrooms, the quizzes are often *ungraded* (Cranney et al., 2009; McDaniel et al., 2007), which may undermine the motivation of students.

We know of only two reports (Hattikudur & Postle, 2011; Khanna, 2015) that have incorporated the testing effect into a graded component of a college-level psychology class. Hattikudur and Postle required students in a cognitive psychology class to complete an online quiz after each day's lecture. However, there was no control condition, so the authors used historical data of how previous students had performed in this class. The quiz-based class did earn a higher final score ($M = 84$) compared to the average from the previous 6 years ($M = 80.3$). This difference was statistically significant, but because data from these classes come from different exams and different cohorts of students, the possibility arises that the tests were easier for the 2010 cohort. Thus, it is possible that if the same test materials were used for all classes, daily quizzes may not have produced higher exam scores.

Recently, Khanna (2015) compared the effects of graded pop quizzes, ungraded pop quizzes, or no quizzes across three Introductory Psychology classes. Khanna reported a statistically significant effect on a cumulative final exam in which students in the ungraded quiz condition ($M = 81.8$) scored significantly higher than students in the graded quiz condition ($M = 75.2$); neither of the quiz groups differed significantly from the no-quiz control class ($M = 78.5$). Khanna interpreted her results in that students in the graded quiz condition experienced heightened anxiety about the pop quizzes, and this anxiety counteracted the benefits of the testing effect. Although this interpretation may be valid, two procedural aspects of Khanna's study are curious and raise the question of whether a different experimental design may yield a better classroom assessment of the testing effect. First, six unannounced quizzes were given during the term to the graded and ungraded conditions, with two quizzes preceding each midterm exam, before the cumulative final exam was administered. Yet, none of the results from questions on these midterm exams are reported. Indeed, completion of these midterm exams should also act as a testing experience, so presumably, even the no-quiz condition received some testing effect benefits. Second, in regard to the one dependent variable that is reported, it is not clear what percentage of the cumulative final exam score was related to questions that appeared on the quizzes and what percentage of the exam indexed other knowledge learned during the term. Thus, the possibility exists that the group differences are unrelated to material experienced during quizzing. Therefore, a more accurate assessment of inducing the testing effect via quizzing would only compare groups on content presented on the quizzes.

The goal of the present study was to examine the testing effect in a naturalistic classroom setting in which the testing effect was assessed by comparing different classes on the same dependent variable. To explore the testing effect in the Introductory Psychology classroom, we conducted a two-group quasi-experiment. The control class represented the study condition in that it was encouraged to read the assigned textbook readings. The quiz class represented the testing effect condition; this class completed brief daily quizzes over the textbook assigned reading. Both groups were later tested on the assigned textbook information on the exam. If the quiz class correctly answered more of these questions than the control class, this would support the hypothesis that the testing effect can be leveraged within an actual college classroom.

Furthermore, to understand better transfer of the testing effect, we devised three different measures. First, some of the test questions were *identical* on the quizzes and the tests; we predicted that the quiz class would do best on these questions, as they were answering them for a second time. Second, some of the test questions were *similar* to the quiz version. These similar questions provided a sensitive measure of the testing effect because although the wording of the question was new, the concept of interest had been tested on an earlier quiz. Third, some of the test questions were *new*, meaning that neither the exact wording nor the key concept tested in the question had appeared on a quiz.

Method

Participants

In this quasi-experimental design, 33 students (15 females, 18 males) were in the control class, which met twice a week on a Tuesday–Thursday schedule (8:30–10:20 a.m.) during fall 2014. There were seven African Americans, two Asian Americans, nine Caucasians, seven Hispanic/Latinos, and eight international/unidentified students. This class was also comprised of 17 first-year students, 10 sophomores, 4 juniors, and 2 visiting international students. The quiz class met 3 times a week on a Monday–Wednesday–Friday schedule (8:30–9:45 a.m.) during winter 2015. In the quiz class, there were 31 students (21 females, 10 males); this class included 5 African Americans, 3 Asian Americans, 16 Caucasians, 4 Hispanics, and 3 international/unidentified students. The quiz class distribution was as follows: 26 first years, 1 sophomore, 1 junior, 1 senior, and 1 visiting international student. On the first day of each class, students completed informed consent forms about their participation in research to improve teaching and learning on our campus and to allow us access to their academic transcripts.

To determine whether the classes were roughly equivalent in academic performance, we obtained high school grade point averages (GPAs) from the registrar's office from 30 students in

each of the classes (some international students did not provide high school GPA in their application materials). The average high school GPA of the control class ($M = 3.7$) and the average high school GPA of the quiz class ($M = 3.8$) did not differ significantly, $t(58) = -.93$, $p = .36$. Thus, even though it was not possible to use random assignment in this quasi-experiment, it appears the groups were equal based on their academic history.

Materials and Procedures

Because the two courses were taught by different instructors using different pedagogical techniques, we first determined a subset of information that was covered in the textbook assigned to students but that would not be covered by either instructor in class. The testing effect was assessed only using the textbook information. Specifically, we identified 19 modules from Myers' *Psychology: Tenth Edition in Modules* (2014), which provided the source material for this experiment. For both classes, the following modules were covered on Test 1 (Modules 11, 12, 13, 14, 15, and 16), Test 2 (Modules 17, 18, 19, 27, 28, 29, 30, and 31), and Test 3 (Modules 32, 33 34, 52, 53, and 54). The average length of these modules was 13.1 pages ($SD = 4.2$, range = 5–21 pages). Importantly, the information from these modules was *not* presented during class meetings; students from both classes were only exposed to this information through reading the text.

The control class was assigned daily readings and they completed three exams. The control class syllabus stated that students would be tested on material presented in class and from assigned readings. On the syllabus, each class meeting had a listed class topic and the assigned reading for that day, which included some readings that would be discussed in class and some that would not (e.g., Thursday, October 23, class topic = memory [Modules 23, 24, 25, and 26]; assigned reading = Module 30, ''Assessing Intelligence''). The control class instructor regularly encouraged the students to keep up with all readings. We considered giving the control class a daily activity to equate with the time spent on quizzes in the quiz class and increase our internal validity, but ultimately, we decided that such a manipulation would likely be perceived as artificial or tangentially related to the course material, so we omitted such a manipulation to preserve the ecological validity of our study.

The quiz class was assigned daily readings, and they completed both daily quizzes on the assigned reading and the three exams. The quiz class syllabus stated that

> based on the extensive data that confirm spaced studying is superior to massed studying for long-term retention, this class will use daily quizzes as a pedagogical technique. Each class will begin with a 5-question, multiple-choice quiz. These questions will be drawn from the previous class meeting and the assigned reading for that day.

Each five-question quiz was composed of two questions from the previous class meeting and three questions taken from Brink's (2013) *Test Bank Volume 2* for the corresponding module from Myers' text. As stated earlier, this module had been assigned as reading for the students (as it had been in the control class), but the material was not covered in class. Each paper-and-pencil quiz was followed by brief feedback during which the instructor read the question and identified the correct answer; no additional explanation was provided. The quizzes were *not* returned to the quiz class students. Each question was worth 1 point, and students completed 21 quizzes (a practice quiz based only on class lecture was given during Week 1 and is not included in the data analyses). Students were allowed to drop their lowest quiz score, so the quizzes totaled 100 points, which was 25% of the final class grade.

Students in both classes completed three exams during the academic term, and each exam contained 15 multiple-choice questions from the test bank for the modules students had been assigned to read (the tests also contained other multiple-choice questions and essay questions unique to each class). Furthermore, these ''module questions'' were categorized as one of the three types: identical question, similar question, or new question. The identical question*s* were those that appeared identically on both a quiz and a test for the quiz class. The similar questions covered a topic that had been presented on a quiz, but the test version was a different question from the test bank over the same core concept. For example, if a Test Bank 2 question on crystallized intelligence appeared on a quiz, a differently constructed question about that concept from the test bank was used on the test. The new questions covered topics from the same assigned module that had not been tested directly on a quiz. For example, a quiz question from the Assessing Intelligence module on Stanford-Binet IQ test was paired with the Flynn effect. Importantly, across the three exams, both classes answered the same 45 module questions. For the control class, these questions were all essentially new, but for the quiz class, these questions were classified as 15 identical questions, 14 similar questions, and 16 new questions. Because we had an unequal number of questions in each category, data were converted to percentage correct for data analyses.

Finally, as a means to determine the quiz classes' evaluation of the quizzes, we constructed a brief five-question exit survey. These questions addressed the students' reading and studying habits. Students answered each question on a 5-point Likert-type scale (1 = *strongly disagree*, 3 = *neutral*, 5 = *strongly agree*). Students completed this survey following completion of their final exam for extra credit.

Results

Figure 1 displays the percentage correct of the identical, similar, and new questions for the control class and the quiz class. Across the three question types, the quiz class correctly answered a higher percentage of questions than the control class. A 2 × 3 mixed analysis of variance (ANOVA) with class as a between factor and question type (identical, similar, and new) as a within factor was conducted. This analysis yielded a significant class effect, $F(1, 62) = 32.9$, $p < .001$, $\eta_p^2 = .347$, as the quiz

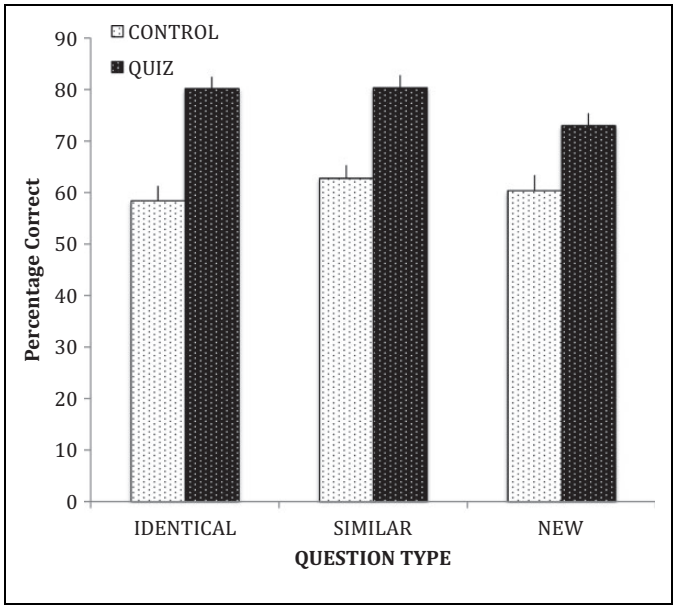

Figure 1. The mean percentage correct on the identical, similar, and new test questions by the control class and the quiz class.

class scored significantly higher than the control class. The question type factor was also significant, $F(2, 124) = 3.5$, $p = .032$, $\eta_p^2 = .054$, but it was qualified by the Significant Class × Question Type interaction, $F(2, 124) = 3.1$, $p = .047$, $\eta_p^2 = .048$.

As a first approach to understand the significant interaction, the two classes were compared on each question type using independent sample *t*-tests. As expected, in regard to the identical questions, the percentage correct of the quiz class ($M = 80.2\%$) was significantly higher than the percentage correct of the control class ($M = 58.4\%$), $t(62) = 5.8$, $p < .001$. The analysis of the similar questions showed a significantly higher percentage correct for the quiz class ($M = 80.4\%$) than the control class ($M = 62.8\%$), $t(62) = 4.2$, $p < .001$. Finally, in regard to the new questions, the percentage correct of the quiz class ($M = 73.1\%$) was significantly higher than the percentage correct of the control class ($M = 60.4\%$), $t(62) = 3.3$, $p = .002$.

As a second means to explore the significant interaction, separate within-group ANOVAs were conducted on the question type data of the control class and the quiz class. No significant question type differences were found in the control class, $F(2, 64) = 1.9$, $p = .153$, $\eta_p^2 = .052$. Yet, a significant question type difference was recorded in the quiz class, $F(2, 60) = 4.3$, $p = .017$, $\eta_p^2 = .126$. Follow-up paired *t*-tests confirmed that the quiz class's performance on identical questions ($M = 80.2\%$) and similar questions ($M = 80.4\%$) was significantly higher than their performance on the new questions ($M = 73.1\%$), $t(30) = 2.4$, $p = .023$ and $t(30) = 3.1$, $p = .004$, respectively. There was no significant difference between the percentage correct of the identical and the similar questions, $t(30) < 1$.

Table 1 displays the mean responses of the quiz class on the five-question exit survey. A one-sample *t*-test was conducted to compare the class responses to the neutral mean of 3. As can be seen in Table 1, all of the comparisons were statistically significant. In sum, the quiz class agreed that the use of daily quizzes helped them change their study habits for the exam; the quizzes increased their reading time, reduced cramming, and redistributed their study time. At the end of the survey, an open-ended prompt invited students to provide other comments regarding the daily quiz requirement. Twenty-five students provided a written comment; only 4 comments were negative, whereas the other 21 were generally positive. An example of a negative comment is the following: "I found the quizzes to be unhelpful, I never knew what was right or wrong." Some representative positive comments include the following: "I feel the quizzes in this class really made me focus on my reading more than just getting the reading out of the way. As much as I hated the quizzes they were a great study technique and made me really focus on the comprehension of my reading"; "I normally hate daily quizzes, but the structure to them in this course made it actually helpful"; and "Keep Doing Them! They held me accountable to get to my winter 8:30 class and helped me study better."

Table 1. Exit Questions From the Quiz Class.

Questions	Mean (*SD*)	*t*-Value (28)
The daily quizzes helped me study for the exam.	4.0 (1.2)	4.4**
The daily quizzes encouraged me to read more in this class than in my other classes.	3.9 (1.1)	5.0**
The daily quizzes reduced the amount of "cramming for the test" compared to other classes.	3.7 (1.3)	3.0*
The daily quizzes required me to distribute my course reading more evenly.	4.4 (0.8)	10.3**
The spacing of the quizzes prompted me to change my study habits in this class.	3.6 (1.1)	3.3*

*$p < .01$. **$p < .001$.

General Discussion

In this quasi-experimental classroom study, the use of daily quizzes improved performance on identical, similar, and new questions compared to a control condition in which students only studied the assigned text modules. We constructed this study as a vehicle to incorporate the testing effect into a college classroom, and the data are consistent with just such an interpretation. Furthermore, the superior performance of the quiz class on the similar and new questions suggests the presence of retrieval-induced facilitation in our study. Students did not have to be tested on the exact same questions they had seen previously in order to benefit; rather, having been quizzed on material seemed to benefit performance on questions about related material as well. As in other studies (cf. Chan, 2010), the presence of retrieval-induced facilitation is noteworthy in this applied setting because it suggests that instructors do not have to "pretest" all information to produce the testing effect benefits. Instead, selective testing of the most pertinent material can strengthen that information and related topics.

The superior performance of the quiz class on the new questions is surprising, in light of a recent hypothesis put forth by Nguyen and McDaniels (2015). In their review of the incorporation of the testing effect into the classroom, Nguyen and McDaniels described the "good, the bad, and the ugly" of using the testing effect. Based on their analysis of lab studies, they suggested the "good effects" of testing would be observed only when the same questions were used during quizzing and testing (our identical condition) or when the quiz and test questions were constructed to examine the same concept (our similar condition). In contrast, they predicted the "ugly effect" of decreased accuracy if test questions were only tangentially related to the quiz questions (as in our new condition). The basis for this prediction from Nguyen and McDaniels arose from their lab-based study in which they compared test and study conditions on similar items and dissimilar items drawn from textbook test bank items. They recorded the expected testing effect on the similar items, but they found the test condition scored significantly lower than the study condition on the dissimilar items. We propose that the discrepancy between the performance of our students and their participants is likely due to the procedural differences between our classroom setting and their lab approach (e.g., the students in our quiz condition had multiple opportunities to study the assigned readings, to study for the quizzes, and to study for the exam). Considering that many instructors like to choose test items from text banks, additional research to explore the effects of transfer on dissimilar quiz test items in a classroom setting is necessary.

Overall, the present results validate earlier work that suggested the testing effect could benefit students in a real classroom (e.g., Butler & Roediger, 2007; Cranney et al., 2009; Hattikudur & Postle, 2011; McDaniel et al., 2007, 2011; Roediger et al., 2011), but they somewhat contradict the findings of Khanna (2015). Khanna reported the students taking graded quizzes did not show a testing effect, and she interpreted this null result as due to increased anxiety in the graded condition. In the present study, the quiz class had daily graded quizzes, and they showed significant testing benefits. The most obvious difference between the two studies is that Khanna employed occasional, unannounced "pop" quizzes, whereas the quiz class students in the present study always knew that a quiz was forthcoming. Therefore, it seems reasonable to conclude that graded, predictable quizzes can improve student performance.

Although we have approached these data from the framework of the testing effect, the fact the quiz class scored significantly higher than the control class on all question types could be due to other factors as well. Instructor or class characteristics, for example, may have played a role because different instructors conducted each class (W.R.B. was the quiz class instructor and J.L.P. was the control class instructor) during different academic terms. Although we have no direct evidence of any such effect, the possibility that students were motivated to read more by the quiz class instructor or during snow-laden winter months cannot be eliminated.

More plausible possibilities are that the daily quizzes promoted more engagement with the material or more time studying in the quiz class than in the control class. Specifically, if the quiz class reviewed the text modules before a quiz and also before the exam, and the control class only read these modules immediately prior to each exam, the additional study time in the quiz class may have produced the class differences. A related possibility is that the daily quiz requirement produced spaced learning in the quiz class whereas the control class may have engaged in massed learning by reading all of the modules immediately prior to the exams. Others have extolled the benefits of frequent quizzing on spaced learning (e.g., Leeming, 2002; Pennebaker, Gosling, & Ferrell, 2013). For example, Pennebaker et al. have recently shown that students who take online daily quizzes get more questions correct than students who answer these same questions on a more traditional "exam." It is notable in the Pennebaker et al. report that students were only tested once (either on a quiz or on a test), so spaced quizzing alone can produce better learning. Currently, the lack of experimental control inherent in the quasi-experimental design prevents us from unequivocally endorsing the testing effect or one of these other factors. Indeed, these various explanations are not mutually exclusive, and they all may contribute to better retention in the quiz class.

Although our focus was to utilize daily quizzes to embed the testing effect within the Introductory Psychology classroom, the explicit use of graded quizzes may not be necessary to produce benefits. A recent meta-analysis of the testing effect by Rowland (2014) suggests that the necessary condition to produce the phenomenon is *effortful recall*. Although graded assignments may provide the best tool to produce motivated effortful recall, many other active learning techniques appear to have a form of effortful recall embedded within them. For example, authors have noted the relative effectiveness of using clickers in the classroom (e.g., Fortner-Wood, Armistead, Marchand, & Morris, 2013; Martyn, 2007). It seems plausible that the requirement for students to attend to the front of the classroom and answer specific questions about a lecture topic resembles the effortful retrieval underlying the testing effect. We postulate that exploration of the common mechanism underlying the testing effect along with other effective teaching techniques may provide fruitful avenues for future laboratory work and pedagogical innovation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors are grateful for financial support from Kalamazoo College's Teaching & Learning Committee.

References

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III, & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, *22*, 861–876.

Brink, J. (2013). *Test bank volume 2 for David G. Myers Psychology tenth edition*. New York, NY: Worth Publishers.

Butler, A. C., & Roediger, H. L., III. (2007). Testing improves long-term retention in a simulated class setting. *European Journal of Cognitive Psychology*, *19*, 514–527.

Carpenter, S. K., & Kelly, J. W. (2012). Tests enhance retention and transfer of spatial learning. *Psychonomic Bulletin & Review*, *19*, 443–448.

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, *14*, 474–478.

Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory*, *18*, 49–57.

Coppens, L. C., Verkoeijen, P. P. J. L., & Rikers, R. M J. P. (2011). Learning Adinkra symbols: The effect of testing. *Journal of Cognitive Psychology*, *23*, 351–357.

Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology*, *21*, 919–940.

Fortner-Wood, C., Armistead, L., Marchand, A., & Morris, F. B. (2013). The effects of student response systems on student learning and attitudes in undergraduate psychology courses. *Teaching of Psychology*, *40*, 26–30.

Hattikudur, S., & Postle, B. R. (2011). Effects of test-enhanced learning in a cognitive psychology class. *Journal of Behavioral and Neuroscience Research*, *9*, 151–157.

Hinze, S. R., & Wiley, J. (2011). Testing the limits of testing effects using completion tests. *Memory*, *19*, 290–304.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, *101*, 621–629.

Kang, S. H. K., McDaniel, M. A., & Pashler, H. (2011). Effects of testing on learning of functions. *Psychonomic Bulletin & Review*, *18*, 998–1005.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528–558.

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968.

Khanna, M. M. (2015). Ungraded pop quizzes: Test-enhanced learning without all the anxiety. *Teaching of Psychology*, *42*, 174–178.

Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology*, *29*, 210–212.

Martyn, M. (2007). Clickers in the classroom: An active learning approach. *Educause Quarterly*, *2*, 71–74.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*, 399–414.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*, 494–513.

Myers, D. G. (2014). *Psychology: Tenth edition in modules*. New York, NY: Worth Publishers.

Nguyen, K., & McDaniel, M. A. (2015). Using quizzing to assist student learning in the classroom: The good, the bad, and the ugly. *Teaching of Psychology*, *42*, 87–92.

Pennebaker, J. W., Gosling, S. D., & Ferrell, J. D. (2013). Daily online testing in large classes: Boosting college performance while reducing achievement gaps. *Plos One*, *8*, 1–6.

Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*, 382–395.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 233–239.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *14*, 1432–1463.